

Aberystwyth University

A NeuroRobotic Model of Infant Looking Behavior

Kerr, Emmett; Triesch, Jochen; Migliore, Davide; Spina, Gabriele; Chandrashekhariah, Pramod; Veale, Richard; Gatsoulis, Yiannis

Publication date:
2011

Citation for published version (APA):

Kerr, E., Triesch, J., Migliore, D., Spina, G., Chandrashekhariah, P., Veale, R., & Gatsoulis, Y. (2011). *A NeuroRobotic Model of Infant Looking Behavior*. Paper presented at Capo Caccia Cognitive Neuromorphic Engineering Workshop, Aberystwyth, United Kingdom of Great Britain and Northern Ireland.
<http://hdl.handle.net/2160/7584>

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

A NeuroRobotic Model of Infant Looking Behavior

Davide Migliore, Emmett Kerr, Gabriele Spina, Pramod Chandrashekhariah,
Richard Veale, Yiannis Gatsoulis, Jochen Triesch

Abstract—Very young human infants demonstrate visual exploration behavior. The behavior is modulated by habituation as stimuli are experienced multiple times. Primate studies have shown that when neural structures responsible for habituation are lesioned, the visual exploration behavior is retained while the habituation (learning) component is abolished. This paper presents an anatomically-inspired neuro-robotic model of the visuomotor (oculomotor) system that can accomplish looking behavior similar to that observed in non-learning infants or in primates with lesioned parahippocampal regions. The neuroanatomical basis for the different parts of the model and their interaction are discussed.

I. INTRODUCTION

Newborn human infants are capable of astounding behaviors with which they are not commonly attributed. For example, visual habituation, a type of visual category learning, is present at least as early as birth [1]. The typical way of testing these (visual) behaviors is to measure the aggregate looking time towards a visual stimulus over multiple encounters with that stimulus [2]. This is compared to the response to a never-experienced visual stimulus that is matched for some properties such as complexity and salience. If the looking time to the “familiar” stimulus is different than the looking time to the “novel” stimulus, then the ability of the infants to discriminate between the stimuli, and even the ability to learn to recognize the stimuli (at least as being “previously encountered vs. not”), is inferred.

In these preferential looking experiments, looking time is a measure aggregated over many discrete behaviors of the infant, such as steady gazes towards a given point (“fixations”), which occur in between shifts between fixations (“saccades”). An infant’s gaze is thus constantly moving at a much faster timescale than the timescale that is usually reported in habituation studies [3]. This micro-level looking behavior (the order, length, etc.) of fixations is interesting because it gives more insight into the mechanisms that are giving rise to the aggregate looking times. They also are informative for building artificial systems to mimic desirable capabilities of the human infants (e.g. developmental learning ability).

Even in the absence of habituation (e.g. with exceedingly simple stimuli such that habituation is instantaneous, or in darkness, or with certain brain areas lesioned) infants and primates demonstrate a visual search behavior. With habituation removed, when aggregated over time, the looking times for stimuli will be proportional to properties of the stimuli such as complexity or salience. It seems there is some intrinsic visuomotor mechanism/dynamics in place running in a constant loop which ensures that gaze will be allocated around the visual environment in a way that is roughly proportional

to (probably low-level) properties of the components of the visual environment. Theories have been suggested (regarding e.g. minimizing uncertainty, entropy, learning rate, etc.) as to “why” this might be the case (evolutionarily), but this is not the question addressed in this paper. This paper rather attempts to build a (robotic) model of what this visuomotor system might look like neurally, and how it functions to produce the looking behavior observed in infants. Habituation’s interaction with the visuomotor system is addressed, but a model of habituation is not explicitly built nor tested in this report. The focus is entirely on the “inside loop” of the system, i.e. the loop which keeps the eyes moving around the visual environment, fixating on components of the environment in turns of lengths roughly proportional to their properties.

The model presented is a “prototype” model containing the “large” pieces of a more complex model of the same phenomenon that is exhaustively based on neuroanatomical evidence. The present model for the sake of prototyping on the robot and demonstration of concept takes some shortcuts anatomically. It does not, however, take shortcuts mechanistically – all mechanisms are accomplished via biologically realistic continuous neural circuit models, and the shortcuts can all be demonstrably implemented using more complex circuit models.

II. TARGET BEHAVIOR

The target behavior for the model will be to match fixation times of very young human infants (less than 2 months postnatal, preferably newborns). The overall looking behavior should be matched qualitatively (i.e. the infant should not only look perfectly between the two most salient objects, it should also look away at e.g. a random point on the wall for some short period of time every now and then). In addition, the proportion of looking time should increase as a function of the stimulus complexity [4]. This function will be simply linear in this model, but as implemented it can be exchanged for any function of the complexity. Thus, it should be matchable to actual infant looking data for e.g. some set of objects. “Stimulus complexity” in this paper and model are just a placeholder for some unknown properties of visual stimuli that draw or hold looking in infants. Particularly, they are placeholders for those properties that experimenters try to control for between stimuli when running habituation or looking time experiments.

III. MODEL DESCRIPTION

Instead of exhaustively enumerating the neuroanatomical, psychophysical, and neuropsychological support for the various

components, the (prototype) model will be laid out, and the salient support presented briefly for the various pieces' functions, maturity at the target postnatal age, connectivity, and role in producing the behavior.

The "architecture" (i.e. implementation-biased description) of the model is presented, rather than the mathematical formalism of the environment and solution. No analysis of the network dynamics is attempted in this paper except for informal observations of the behavior of the network in the robotic infant while it is "behaving".

The network is made up of several main pieces (Figure 1). The overall theory of its functionality is illustrated in Figure 2.

A. Saliency Map

Images (frames) come in (actually, one from each eye, left and right) and are processed by a saliency map ([5]), which assigns a salience value to each location in the image by looking for areas of high global "uniqueness" at many spatial scales. This is accomplished by applying filters at many scales in parallel to detect feature channels such as motion, color, intensity, orientation, etc., and then subtracting and subsampling between sequential spatial scales, within feature channels. This is meant to represent some sort of "fast scene processing" salience, which does not care so much about the content at visual locations so much as the "salience" of locations in relation to the rest of the visual field. It is imagined that this type of saliency map would be at play in orienting behavior, which is what the looking behavior in this paper is intended to represent. Some of the saliency channels (color, orientation) are not ones that would be processed by the superficial layers of the superior colliculus (SC)(mostly responds to movement), nor is there evidence for (explicit) spatial scales even in V1, so it is not clear where a salience map would actually be implemented, especially in an infant. Some theories have been offered arguing that horizontal competition within V1 could result in a saliency-map-like phenomenon [6], but these are outside the intended scope of the model presented in this report.

B. Superior Colliculus - deep (integration, eye control)

The instantaneous salience map calculated for every frame is treated as "input" into a longer-term map made up of leaky integrating neurons (non-firing). This is considered to correspond to the intermediate/deep layers of SC, especially since this is the last "stop" before eye movement signals in the model. Each neuron's membrane potential V_m is described by the equation:

$$\frac{\partial V_m}{\partial t} = \frac{-(V_m - V_{rest}) + R_m \cdot (I_{bg} + I_{syn})}{\tau_m} \quad (1)$$

where R_m is the membrane resistance (uniformly 1.0 M Ω for all neurons), I_{bg} the background current (uniformly 0.0 nA for all neurons) and I_{syn} the total current impinging from afferent synapses. The $-V_m$ represents the leakage term, causing the membrane potential to decay exponentially with time constant τ_m (uniformly 30 ms for all neurons, though since update

was per-frame this has no relation to real-time). The resting potential of the membrane V_{rest} is assumed to be 0 mV.

Afferent input into a dSC neuron from the instantaneous saliency map is simply linearly summed into I_{syn} , after being multiplied by the efficacy of the synapse connecting it. The synaptic weights and connections are built initially using a 2-dimensional gaussian, with standard deviation ω_w (1.5 for the experiments) and a cutoff value (i.e. minimum weight outside of which region there are no connections) χ (0.05 for experiments). This has the effect of slightly blurring (low-pass-filtering) the instantaneous salmap (thus bringing out "masses" of high salience more than just isolated pixels of high salience). It also mimics the increase in receptive field size one sees when moving more towards the dSC (many pre-synaptic neurons connect to fewer post-synaptic neurons, in retinotopically defined areas).

C. Reticular Formation (arousal response to complexity)

Physiological functions (arousal, measured by heart rate) have been shown to be correlated with attentional phases in infants [7]. Some have hypothesized the underlying circuits for these include e.g. the reticular formation (RF) which is responsible for the ascending transport of neuromodulators (hormones) to diverse regions of the brain [8]. Such areas could become more active in response to stimulating environments (i.e. with more complex stimuli, more variation, etc.) and are also responsible for the lack of responsiveness of infants without sufficient stimulation. It's important to note that endogenous factors guide a majority portion of the very young infant's arousal/attentional state at any time. It's difficult to get very young infants to respond to exogenous stimuli if they are not already in an "alert" state. This report/model does attempt to deal with these factors either.

The complexity-responsive functionality of the RF circuits are used as a mechanism for determining (soft) looking time towards an object. Really, it could be implemented via other methods such as simply probabilistic microsaccades that are more likely to get "caught" locally in the more complex stimulus. Since usually the amount of looking time towards a stimulus during e.g. habituation experiments is postulated to be a function of "encoding" of the stimulus (and thus, longer for more complex stimuli), yet we have explicitly lesioned those areas that are "making progress" in any sense (or learning), there must be some other mechanism than simply "learning" that drives looking time towards stimuli (though when present, feedback from those learning regions certainly may play a role). Since in habituation experiments the salience etc. of stimuli is controlled, it is not trivial to decouple the habituation response from the "baseline" stimulus response. Note also that more complex mechanisms, e.g. encoding into some sort of short-term memory, waiting for a non-transient response/cycle of some minimal error threshold in response to the stimulus, could also be how this stimulus complexity causes longer looking times. The current mechanism at least attempts to capture the phenomenon, if not the mechanism. This is perhaps one of the least understood areas of oculomotor behavior –

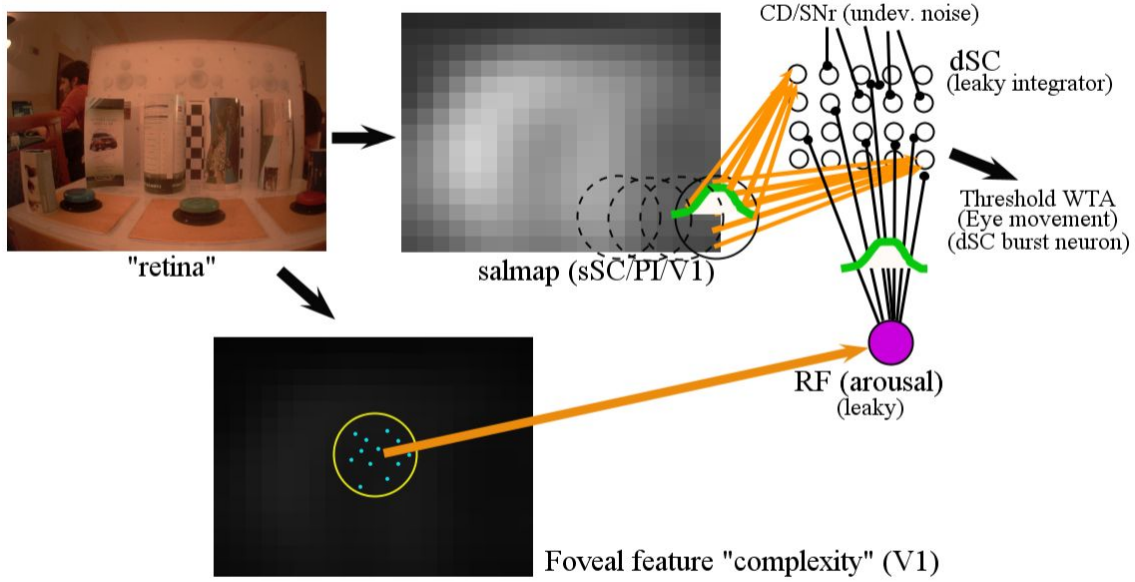


Fig. 1. Graphical overview of connectivity and regions of network, and the general function of each region. Full-model anatomical correlates listed next to regions. Links between regions are not intended to imply direct anatomical connections since the model has been simplified.

namely, in the absence of any scene change, what causes the infant to look away? Several observations exist that correlate with the phenomenon, e.g. the dying out of activity in fixation neurons in dSC [9], changes in activity in SNr [10], CD, and the buildup of burst neurons in dSC [11], but the real dynamics and mechanisms and how all these work together to move attention (fixation) are not well understood. This is one of the things this model and, later, the more neurally accurate one hope to offer explanations for.

In the model the RF is modeled as a single leaky neuron. One can think of it as population coding of the activity of some region that got a shot of neuromodulator/hormone cocktail in response to exciting foveal stimuli. The RF neuron inhibits every region of the dSC, so that more activation of the RF will result in more inhibition of SC (pushing the neurons there to very negative potentials), and thus it will take longer for SC neurons to recover via the normal continuous input coming from e.g. the instantaneous saliency map and the gaussian noise from SNr. However, to prevent the continuous fixation over multiple “refixations” of the central area (which might not be coded in mammals, i.e. we wouldn’t have to worry about the currently foveated area ever getting burst neurons activated since there are no corresponding burst neurons there) the weighting of this inhibitory projection is topologically modulated, such that more foveal/central positions have stronger inhibitory weights than the periphery (which still have some baseline inhibitory weight).

Thus, the weight of a point in the retinotopic projection as a function of distance from the centre of the fovea is equal to the value at that point of a gaussian centred on the centre, with amplitude α (6.0), variance σ (1.5), and baseline (i.e. y-shift) β (1.0).

When a gaze shift is made, the “complexity” c_f of the

current foveal content is scaled and injected into the RF neuron (the RF neuron’s V_m is set to c_f). In this case, as a stand-in, the complexity is calculated as simply the number of Harris feature points (thresholded 1-d gaussians LPFs run along the x- and y-dimensions) within a radius r of the centre of the image. In binocular cases the two eyes are simply summed together linearly.

The RF neuron’s membrane potential (V_{RF}) decays exponentially with some time constant τ_{RF} (500ms, though again since in the experiments update of the saliency model is not locked to real time but rather to the refresh speed of the cameras, the units are arbitrary w.r.t. the experiments). At any rate, the equation stands:

$$\frac{\partial V_{RF}}{\partial t} = \frac{-V_{RF}}{\tau_m} \quad (2)$$

The membrane potential of RF is injected directly into all (e.g. dSC) post-synaptic neurons, linearly scaled by the weight of the synapse connecting RF with that neuron.

D. Substantia Nigra pars reticulata (Gaussian noise)

Caudate Nucleus (CD) and SNr neurons show visual and habituation responses [10], which implies they are at play in normal visual search behavior and/or habituation learning. Assuming it’s just habituation or even considering they are engaged in disengagement and thus normal visual looking behavior ([12]), these disengagement and habituation behaviors are not seen in infants or in limbic-cortex lesioned primates, respectively [13]. This suggests that the SNr itself, it’s afferents, or its efferents to dSC are undeveloped in the young infants. SNr-dSC projections are known to be strongly GABAergic and thus inhibitory, and to receive inhibitory projections from CD (“dis-inhibitory”, which in turns gets input from a variety of cortical regions as well as having interesting intrinsic dynamics

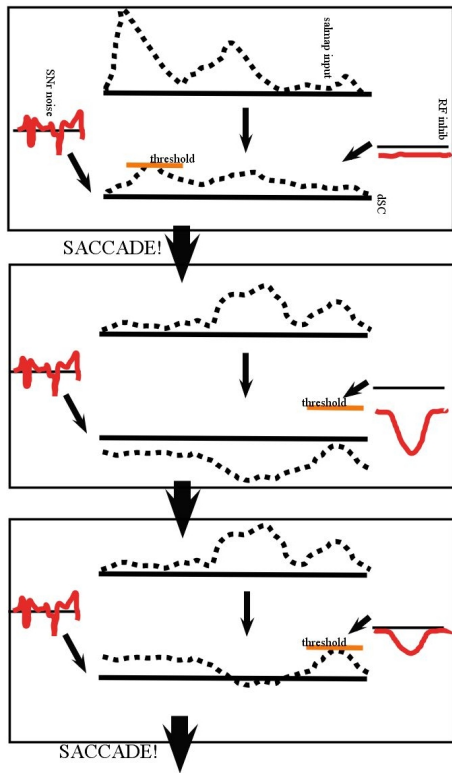


Fig. 2. Intended loop dynamics of the model through several fixation-saccade iterations. Intended local circuit response properties listed where predictable.

involving dopamine via SNr pars compacta (SNc)). We model these dynamics and the product of immaturity of this area or its dSC efferents by simply considering the baseline of SC as already receiving some amount of inhibition from the tonically firing SNr GABAergic neurons, and then add or subtract gaussian noise (gaussian random variables) from the synaptic input to each dSC neuron on every time step. The probability of drawing an input of strength of δ is based on the mass of the gaussian at the point, thus small numbers are likely, and very extreme numbers are less likely. In the experiments, the noise is drawn from a gaussian with σ of 0.01 and amplitude (scale of output inputs) of 1.0.

IV. IMPLEMENTATION ON ICUB ROBOTIC PLATFORM

The prototype model was implemented on an iCub robot head (Figure 3). It has two pan-tilt-vergence eyes mounted in the head supported by a yaw-pitch-twist neck. It has 6 degrees of freedom (3 for the neck and 3 for the cameras). We make use of stereo information to select the most salient point in the scene. The saliency maps from both the eyes are input to the decision module to select the single most salient point in the scene irrespective of which eye the saliency point comes from (intended to model the result of contralateral connections between the two lateral dSC, and the known diffuse inhibitory input from the contralateral SNr [14]). However, the saliency points only at the interest points in the scene are considered for selecting a target for a saccade.

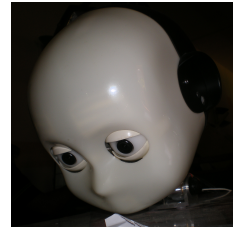


Fig. 3. The iCub head that was used for the experiments.

A. Interest points add stability to the scene information

An interest point is a point in the image which has a well-defined position in image space that is rich in terms of local information content around it and is stable against the variations such as brightness, scale etc so that it is repeatable from frame to frame. It is not clear as to how much the ability to locate such interest points is pronounced in infants, however, studies show that V1 areas of visual cortex are sensitive to many simple features such as orientations at a very early age [15]. At the very least there exists a mechanism to identify discriminant features in growing infants, since they without fail begin to identify and distinguish objects at some point.

In this work, we use Harris corner points as interest points. A corner point is defined as the intersection of two edges or as a point for which there are two dominant and different edge directions in a local neighborhood of the point. Corner points serves the characteristic of an interest point as it has a well-defined position and can be robustly detected. To determine a Harris corner point the autocorrelation matrix of the second derivative of images over a small window around each point is calculated and the two largest eigenvalues are tested to determine whether they are greater than a threshold. Only those points that pass the threshold are considered.

B. Choosing the attention point in stereo

Once the Harris corner points are calculated on the left and right images, we attempt to find corresponding points in the left and right images. Those points that do not have a matching point in the other image are removed. We then evaluate the value of saliency at the regions containing the remaining harris corner points on both the images. The weight distribution from the instantaneous saliency map to the long-term saliency map implements a gaussian kernel, and thus the dSC values for those points in the image represent an average response of the saliency map around the harris points. We select the most salient point from these values and find its corresponding location on the other eye by its best matched feature. This location in the stereo vision is where iCub's eyes should be directed (see figure 4).

C. Vergence control Mechanism

The iCub has to move its gaze towards the salient point for the object to be focused in the center of the camera view. This task is accomplished using a tracking scheme that is completely object model free (Democratic Cue Integration [16]).

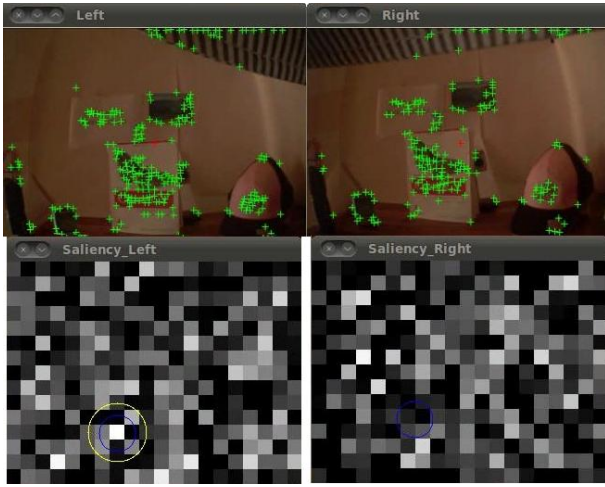


Fig. 4. Saliency values at the Harris corner points are calculated and the maximum is used to steer the eyes. The blue circle indicates the maximum salient point in each eye and the yellow circle indicates the winning point with the highest saliency between the eyes.

Democratic cue integration is a multi-cue tracking system that provides a fast and robust way to track unknown objects in a changing environment (lighting, background, noise, object motion). It combines information from various cues each with a weighting coefficient that depends on their agreement with all other cues. This makes the system adaptive to a given situation as the cues that are not suitable quickly lose prominence. In our work we use motion, kalman prediction, template matching, color kernel tracking and contrast based kernel as different cues.

Tracking executed in the above manner provides a very robust coordinated movement of the head and eyes, which is usually a challenging problem especially in binocular vision. One can observe that the movements accomplished by this method of a vergence-control run in a closed loop with visual input. It is not clear how the problem of vergence is solved in infants but it seems to be clear that saccades are not closed loop or vision-driven but are rather ballistic and motor-driven [17]. Since the saliency model was “turned off” during eye movements, this had no effect on the collected data, though from a real-time point of view it was somewhat slow compared with actual saccades (limitations of the motors and the algorithms for vergence). Coordinated movement of the head and eyes is likewise probably unrealistic in such young infants.

D. Segmenting out the target object from the salient location

For the next step, we use stereo disparity information to coarsely segment the focussed object from the background. We calculate the stereo disparity at all interest points in the scene. To do this we calculate Gabor-jets at interest points on both left and right images. We then compare jets at all interest points in the left image with all points on the right image and we choose the best matching pairs of points by comparing the similarity between the jets (inner product between the

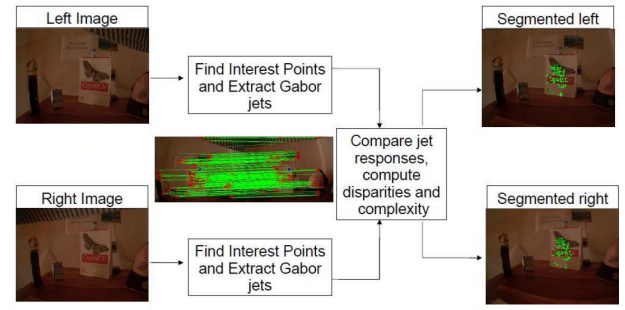


Fig. 5. Stereo segmentation by matching Gabor-jets at the interest points on left and right camera images.



Fig. 6. A picture of the experimental setup that was used during the experiments. The type of environment presented can be seen in front of the iCub head in the upper-left corner.

vectors should be above a threshold value set to 0.93). We then calculate the stereo disparity (horizontal as well as vertical) for each matching pair. The disparity of the interest points falling on the object that has been focussed is considered as the reference value and the points that are away from this value (more than 2 in both +ve and -ve directions) are considered to be from the other objects or background and hence removed. This procedure gives us a rough segmentation of the object or at least segregates the interest points (and hence the feature points i.e. gabor jets) belonging to the object from the rest (see figure 5). This would be helpful for initiating a learning process of the object for future work that integrates learning with the looking behavior which was the focus in this report.

E. Fixation guided by complexity

When the robot looks at a point, the time spent in looking at the point (usually an “object”) should be dependent on the complexity of the point (in our experiment, it is made dependent on the number of Harris corner points as already discussed in the previous section). The trajectory (i.e. ordering) of looking among the objects is recorded.

Based on runs on single frames (see `testonppm.cpp`) the decaying-inhibition dynamics of the model work as intended. To demonstrate this, see Figure 7, plotting the results of automated experiments with the model measuring the time

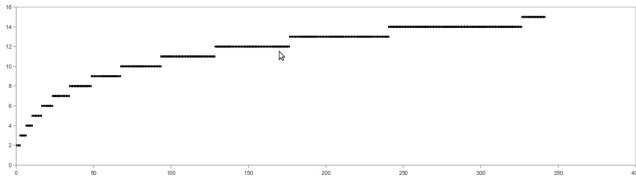


Fig. 7. Plot of the time between fixations (i.e. fixation time to that foveal region) for varying levels of foveal complexity (the manipulated variable in these experiments). Note the logarithmic pattern that emerges: for very low complexities, fixation is disengaged quickly, and for slightly different complexities the looking time differs, but as complexity increases, the looking time saturates to a maximum.

between first fixation and second fixation as a function of the complexity of the foveal image (x-axis).

This is the desired behavior, and would result in (at least) fixation times that follow the desired model of fixation time. For the longer-term looking dynamics, however, since foveation of the salient stimulus is assumed (which entails a change in the visual field since the eyes/head move), it's impossible to test the looking dynamics without running on the robot or in a simulation environment that moves the visual field.

V. AN EXPERIMENTAL BOARD FOR LEARNING SEQUENCES OF ACTIONS BASED ON VISUAL REWARDS

One aspect of the extensive work described in this report was to give to encourage the robot to learn particular sequence of gazes/actions.

This would be achieved by using an experimental board which has been designed and developed by the IMCLeVeR consortium for this particular purpose, i.e. to be able to pre-program sequences of actions that lead to rewards. The board, which is shown in Figure 8, consists of two main types of modules:

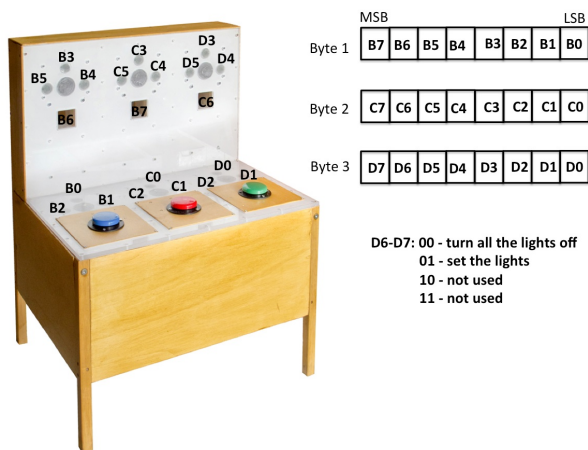


Fig. 8. Experimental board

- Actions modules which the agent interacts with
 - These are buttons () that can be pressed down

- Rewards modules that are activated according to pre-programmed sequences of the actions modules
 - Lights that switch on/off
 - Small cabinets with lights that can open up
 - Audio feedback

More details about the board can be found in the IMCLeVeR project website¹.

The current software interface of the board is written in Labview, and allows the activation of any of the rewards for pre-programmed sequences of actions. Although this control interface provides a quick way of pre-programming the behaviour of the board in a user-friendly way, it has the main limitation that it does not allow dynamic online re-configuration of the board from the controller program of the agent. In our case there was an additional motivation for requiring dynamic control of the board from the controller of the agent; as the available robot platform consisted of only an iCub head, the lack of any actuators prohibited any interaction with the actions modules. For this reason, it was decided to devise a scheme of “telekinesis”, in which the actions modules could be pushed when the iCub focussed its gaze on them and certain rewards would be given, e.g. lights switch on. As such, we implemented an interface that provides dynamic and online control of the lights rewards through the serial port, and can be used by any controller.

A. Technical

On the board there is a usb hub which connects a number of I/O usb cards controlling the lights, the buttons, the motors of the rewards draws, the audio, etc. For the serial control interface the I/O cards are connected directly to the usb ports of a PC. The usb-to-serial converter may require a driver which can be found in the manufacturer’s website², although no installation was necessary as the driver seems to be included in the linux kernel 2.6.32-31, and as such I assume in any more recent ones.

The serial control interface was written in python (2.6.5), and the development platform was running Ubuntu Linux 10.04 64-bit with kernel 2.6.32-31-generic.

The board we had available had a slightly different wiring from the original one shown in Figure 9. Its differences are highlighted in Figure 9.

For the control of the lights, the writing of a 3 – byte message to the (usb-to-)serial port which is connected to the PC is needed. The order of the bits are shown in Figure 9. The main file is `board_serial_control.py`, which provides the class `C_BoardSerialControl` that consists of `C_LightsSerialControl` that is responsible for controlling the lights. If the wiring of another board is different from the one provided, all is needed is to change the mapping dictionary (line 60) to the correct mapping. A small utility (`board_quick_test_serial_lights.py`) can be used to identify the wiring of your board. Also the programs

¹<http://www.im-clever.eu/>

²<http://www.ftdichip.com/Drivers/VCP.htm>

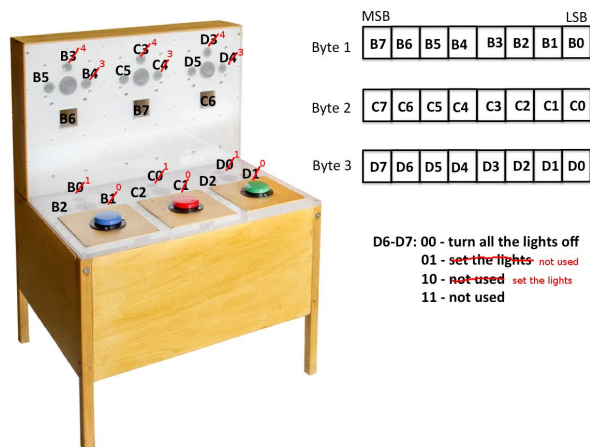


Fig. 9. Experimental board that was available

demo.py and demo_sections.py demonstrate the use of the provided interface. Lastly, the interface can be easily extended for the rest of the modules on the board if they follow a similar interface.

VI. CONCLUSION

By design, the mechanism to determine looking time towards a foveal stimulus as a function of its complexity was successful, but this is not the interesting result since these dynamics are predicted and necessitated by the simple mechanism.

The interesting result will be the inter-object looking behavior in a real environment. Will, and how often will, the system make saccades to “noise” areas but then quickly look away because of lack of complexity?

In the future, the model will be implemented as the more complex anatomical circuit, the saliency map replaced with proper neural code from the eye-up, and the actual neurons “fudged” in most of the areas and their dynamics actually simulated. The benefit of building the more realistic model will be to actually look at the different dynamics of the model in reaction to the changing environment, since that kind of interaction can not be easily predicted or simulated.

In addition, in the future we intend to explore the interaction of a learning system (i.e. a habituation mechanism) and how it interacts with the simple ocular-motor control system implemented in this report.

REFERENCES

- [1] A. Slater, V. Morison, and D. Rose, “Habituation in the newborn,” *Infant behavior and development*, vol. 7, pp. 183–200, 1984.
- [2] R. Fantz, “Visual experience in infants: Decreased attention familiar patterns relative to novel ones,” *Science*, vol. 146, pp. 668–670, 1964.
- [3] G. W. Bronson, “Changes in infants’ visual scanning across the 2- to 14-week age period,” *Journal of Experimental Child Psychology*, vol. 49, pp. 101–125, 1990.
- [4] M. Lewis, J. Kagan, and J. Kalafat, “Patterns of fixation in the young infant,” *Child Development*, vol. 37, no. 2, pp. 331–341, 1966.
- [5] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

- [6] Z. Li, “A saliency map in primary visual cortex,” *Trends in Cognitive Sciences*, vol. 6, no. 1, pp. 9–16, 2002.
- [7] J. E. Richards and B. J. Casey, “Infant visual recognition memory performance as a function of heart rate defined phases of attention,” *Infant Behavior and Development*, vol. 13, p. 585, 1990.
- [8] R. W. Doty, “Brainstem influences on forebrain processes, including memory,” in *Neurobehavioral plasticity: learning, development, and response to brain insults*, N. E. Spear, L. P. Spear, M. L. Woodruff, and R. L. Isaacson, Eds. Erlbaum, 1995, pp. 349–370.
- [9] D. Munoz and R. Wurtz, “Saccade-related activity in monkey superior colliculus. i. characteristics of burst and buildup cells,” *Journal of Neurophysiology*, vol. 73, no. 6, p. 2313, 1995.
- [10] O. Hikosaka and R. H. Wurtz, “Visual and oculomotor functions of monkey substantia nigra pars reticulata. i. relation of visual and auditory responses to saccades,” *Journal of Neurophysiology*, vol. 49, no. 5, pp. 1230–1253, 1983.
- [11] D. Munoz and R. Wurtz, “Saccade-related activity in monkey superior colliculus. ii. spread of activity during saccades,” *Journal of Neurophysiology*, vol. 73, no. 6, p. 2334, 1995.
- [12] M. H. Johnson, “Cortical maturation and the development of visual attention in early infancy,” *J. Cognitive Neuroscience*, vol. 2, no. 2, pp. 81–95, 1990.
- [13] J. Bachevalier, M. Brickson, and C. Hagger, “Limbic-dependent recognition memory in monkeys develops early in infancy,” *NeuroReport*, vol. 4, pp. 77–80, 1993.
- [14] H. Jiang, B. Stein, and J. McHaffie, “Opposing basal ganglia processes shape midbrain visuomotor activity bilaterally,” *Nature*, vol. 423, pp. 982–986, 2003.
- [15] C. Von Der Malsburg, “Self-organization of orientation sensitive cells in the striate cortex,” *Kybernetik*, vol. 14, pp. 85–100, 1973.
- [16] J. Triesch and C. Von Der Malsburg, “Democratic integration: Self-organized integration of adaptive cues,” *Neural Comput.*, vol. 13, pp. 2049–2074, September 2001. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1119715.1119722>
- [17] L. Hainline, “Summary and commentary: Eye movements, attention, and development,” in *Cognitive Neuroscience of Attention: A developmental perspective*, J. E. Richards, Ed. Lawrence Erlbaum, 1998, pp. 163–178.